Confirmatory analysis of exploratively obtained factor structures

van Prooijen, J.W.; van der Kloot, W.A.

published in
Education and Psychological Measurement
2001

DOI (link to publisher)
10.1177/00131640121971518

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
van Prooijen, J. W., & van der Kloot, W. A. (2001). Confirmatory analysis of exploratively obtained factor structures. *Education and Psychological Measurement*, *61*(5), 777-792. https://doi.org/10.1177/00131640121971518

Educational and Psychological Measurement
http://epm.sagepub.com/

Confirmatory Analysis of Exploratively Obtained Factor Structures

Jan-Willem van Prooijen and Willem A. van der Kloot

Educational and Psychological Measurement 2001 61: 777
DOI: 10.1177/00131640121971518

The online version of this article can be found at:
http://epm.sagepub.com/content/61/5/777

Published by:
SAGE
http://www.sagepublications.com

CONFIRMATORY ANALYSIS OF EXPLORATIVELY OBTAINED FACTOR STRUCTURES

JAN-WILLEM VAN PROOIJEN AND WILLEM A. VAN DER KLOOT
Leiden University

Factor structures obtained by exploratory factor analysis (EFA) often turn out to fit poorly in confirmative follow-up studies. In the present study, the authors assessed the extent to which results obtained in EFA studies can be replicated by confirmatory factor analysis (CFA) in the same sample. More specifically, the authors used CFA to test three different factor models on several correlation matrices of exploratively obtained factor structures that were reported in the literature. The factor models varied with respect to the role of the smaller factor pattern coefficients. Results showed that confirmatory factor models in which all low EFA pattern coefficients were fixed to zero fitted especially poorly. The authors conclude that it may be justified to use a less constrained model when testing a factor model by allowing some correlation among the factors and some of the lower factor pattern coefficients to differ from zero.

Multifactor psychological scales and tests are very common measurement tools that are widely used in research and in practical applications. Psychological knowledge is, for a large part, founded on such psychometric instruments. Most multifactor tests and measurement instruments are initially developed by exploratory factor analysis (EFA) (Gorsuch, 1983), which produces a set of factor pattern coefficients on, usually, orthogonal factors based on the correlations between the items of the test. If one wants to investigate whether the factor structure of the test can be replicated in a new study, one could, in principle, perform EFA on the new data. Subsequently, one

The authors would like to thank Dr. Ab Mooijaart, Ilja van Beest, Kate Hudson, Kees van den Bos, Stephen G. West, and two anonymous reviewers for their valuable comments on previous drafts of this article. Correspondence concerning this article should be addressed to Jan-Willem van Prooijen, who is now at the Free University Amsterdam, Department of Social Psychology, Van der Boechorststraat 1, 1081 BT Amsterdam, the Netherlands; e-mail: jw.van.prooijen@psy.vu.nl.

Educational and Psychological Measurement, Vol. 61 No. 5, October 2001 777-792

could check whether the number of factors, the amount of variance explained, the factor pattern coefficients and factor correlations, as well as the corresponding interpretation corroborate the original results. However, as EFA is data driven and involves a number of subjective decisions, the more appropriate way to cross validate the factor structure of a test is by means of confirmatory factor analysis (CFA) (Byrne, 1989; Jöreskog & Sörbom, 1989; Pedhazur & Schmelkin, 1991). The basic question answered by CFA is whether the factor structure matches the results of the original study.

More specifically, in CFA, a model is tested that specifies in advance the relations between observed variables and latent factors and the relationship among the factors themselves. Such a model contains parameters that are (a) fixed to a certain value, (b) constrained to be equal to other parameters, and (c) free to take on any unknown value. The fit of the model to the data can be statistically evaluated by means of a χ^2 statistic. The χ^2 test, however, is very sensitive to conceptually unrelated technical conditions, like sample size (Bandalos, 1993; Boomsma, 1987) or a violation of the multivariate normality assumption (Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992; West, Finch, & Curran, 1995). Therefore, the fit of the model is often evaluated by means of a group of descriptive fit indices, such as the Bentler-Bonett nonnormed fit index (NNFI) and the comparative fit index (CFI) (e.g., Bentler, 1990; Bentler & Bonett, 1980; Bollen, 1986, 1990; Hu & Bentler, 1998; Marsh, Balla, & McDonald, 1988; McDonald & Marsh, 1990).

When using CFA in practice, a number of decisions have to be made with regard to the model that one should test. Typically, the interpretation of exploratively obtained factor structures involves some simplification. Factors are distinguished on the basis of high factor pattern coefficients. Items with lower pattern coefficients (usually, pattern coefficients with $|f_{ij}| < .3$) are ignored, both with regard to interpretation of the factors and for combining items into composite measures of separate factors (cf. Grice & Harris, 1998). This practice often leads to a dilemma for CFA. Should one test a factor model in which all factor pattern coefficients are specified, including the low ones? Or should one try to confirm a simpler model in which the lower pattern coefficients are fixed to zero?

The research literature shows many instances in which the claimed factor structure of a set of variables (e.g., subscales of a measurement instrument) could not be confirmed by CFA in a subsequent study. In particular, there are many examples of exploratively obtained factor structures that could not be replicated in subsequent CFA studies (e.g., see Borkenau & Ostendorf, 1990; Church & Burke, 1994; Hartman et al., 1999; Lonigan, Hooe, David, & Kistner, 1999; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996; Parker, Bagby, & Summerfeldt, 1993; Rao & Sachs, 1999; Reynolds & Lee, 1991; Vassend & Skrondal, 1997). For these discrepancies, several explanations are possible, both substantive and methodological. Substantive explanations include, for example, that the original measurement structure does not apply to the present participant population, or the measurement structure has been changed through translation (cf. Van de Vijver & Leung, 1997). Researchers may sometimes be tempted to adopt such a substantive explanation. For example, in cross-cultural studies, researchers may attribute differences between EFA and CFA solutions to cultural differences between populations (e.g., Rao & Sachs, 1999). Furthermore, researchers may dispute the validity or generalizability of a scale after the failure to confirm its original factor structure (Hartman et al., 1999; Parker et al., 1993). Although substantive explanations like these may very well be appropriate explanations for a lack of fit in CFA, we argue here that methodological explanations should be ruled out first. After all, substantive explanations in cases in which a lack of fit in CFA should be explained methodologically may contain the risk of drawing wrong conclusions, which may lead to erroneous psychological theory building. We will now discuss three classes of methodological explanations for apparent discrepancies in results of EFA and CFA.

One possible explanation for a lack of correspondence between EFA and CFA results is that many exploratory factor solutions that are described in the literature tend to be based on inadequate applications of EFA, which, in turn, may have led to wrong factor solutions. As Fabrigar, Wegener, MacCallum, and Strahan (1999) have convincingly demonstrated, many EFAs reported in the literature were based on (a) inappropriate criteria to determine the number of factors, (b) an inappropriate rotation method, and (c) an inappropriate factor analytic procedure (notably, principal component analysis [PCA] instead of EFA). These inadequacies were shown to have important consequences for the derived factor structures. For example, the exploratory analysis may not have discovered the appropriate number of factors. Therefore, we conclude that it is highly likely that researchers sometimes fail to confirm a factor model because the original model did not have a solid empirical basis in the first place.

Another possible explanation for a lack of correspondence between results from EFA and CFA is that the two techniques may not be fully comparable. That is, EFA is primarily a data-driven technique and therefore allows researchers much freedom with regard to the number of factors one wishes to retain in the model. Furthermore, in EFA all variables incorporated in the analysis are free to load on all factors (e.g., Gorsuch, 1983). Conversely, CFA is a theory-driven rather than a data-driven technique (e.g., Bollen, 1989). That is, researchers have to specify the number of factors in advance. In addition, not all variables are free to load on all factors: the factor pattern coefficients that were "low" or "insignificant" in the original exploratory analysis (usually, pattern coefficients with $|f_{ij}| < .30$) are often fixed to zero in the confirmatory analysis. Therefore, a number of parameters that were unconstrained in the exploratory analysis are restricted in the subsequent confirmatory analysis. Because CFA typically has more restrictions than EFA, it is therefore by nature more conservative than EFA (cf. Bollen, 1989).

Insofar as this difference in conservativeness between EFA and CFA is responsible for the lack of fit of an EFA-based model in a CFA analysis, it could mean either of two things. On one hand, the lack of fit of CFA seems to suggest that CFA may sometimes be too conservative, and relatively small and unimportant deviations from the model often lead to model rejection. On the other hand, the lack of fit may also suggest that, sometimes, EFA is too liberal, because it may be too easy to interpret an exploratory factor solution as satisfactory, whereas the "true" model may be different from the one originally retained.[1] In fact, this latter position is implied by the criticism by Fabrigar et al. (1999) on the injudicious use of some forms of EFA.

Finally, some divergences in results obtained by EFA and CFA may be explained by inappropriate applications of CFA. That is, to obtain a good fit, a confirmatory factor model sometimes has to be adjusted in several ways. Such modifications of the original model do, however, abandon the theory-driven confirmatory logic of CFA and may make the results almost as data driven as EFA. Although adjustments of the confirmatory factor model may increase the fit of the model, it may at the same time lead to a model that is substantially different from the theoretical model originally hypothesized on the basis of an earlier exploratory analysis.

In a typical cross-validation study, CFA is carried out on new data collected independently of the original study in which a factor structure was derived by EFA. If the two techniques are applied to different data sets, it is difficult to ascertain whether the three methodological explanations mentioned above (i.e., inappropriate applications of EFA, incomparability of EFA and CFA, and inappropriate applications of CFA) accounts for a lack of correspondence between EFA and CFA. After all, it is easy to imagine a substantive reason why CFA failed to replicate results of EFA on different data (e.g., different participant populations or the variables have different meanings in these populations; cf. Van de Vijver & Leung, 1997).

To rule out such substantive explanations, the ultimate test of possible methodological explanations for the differences in EFA and CFA results should necessarily involve the *same* samples. That is, the same data set is used to derive a factor model by EFA and subsequently test this model by CFA. From a substantive point of view, EFA and CFA should lead to the same conclusions when applied to the same data. Therefore, only methodological explanations can account for cases in which EFA and CFA lead to different conclusions based on the same data. This is important because if CFA cannot confirm results of EFA on the same data, one cannot expect that CFA will confirm results of EFA in a different sample or population. To summarize, to judge the (lack of) fit of CFA on new data, it would be useful to know more about the (lack of) fit of CFA on the same data from which the factor model was derived.

The Present Research

In the present study, we investigated the extent to which results obtained in EFA studies can be replicated by CFA on the same set of observations. Furthermore, we studied whether the aforementioned methodological explanations can account for possible discrepancies between results obtained by EFA and CFA. We thus aimed to understand the extent to which applied researchers can expect that results obtained in EFA will replicate in a confirmatory analysis. To these ends, we reanalyzed several correlation matrices published in the research literature and systematically performed confirmative factor analyses on the *original data* for which EFA had been conducted and published. To address the three methodological explanations for differences between EFA and CFA (i.e., inappropriate applications of EFA, incomparability of EFA and CFA, and inappropriate applications of CFA), we tested three different confirmatory analysis models on each correlation matrix.

In the first model (Model 1), we tested the *exact* factor structure obtained by the EFA. That is, all of the pattern coefficients (i.e., both high and low factor pattern coefficients) were fixed to their original values as obtained in the original EFA and, when applicable, the correlations among the factors obtained in EFA (e.g., after an oblimin rotation in EFA) were also fixed. This "baseline" model allowed us to perform CFA on the exact factor structure in EFA on the same data and thus provides a confirmative test of this factor structure. In doing so, we evaluated the extent to which the original exploratory model would be suitable for confirmation. That is, a bad fit of this model may suggest that the EFA did not discover the optimal number of factors or the most appropriate factor pattern coefficients. On the basis of the study by Fabrigar et al. (1999), we expected that a confirmatory test of Model 1 would not fit optimally in all cases, even though this test uses the exact factor structure obtained in EFA from the very same data.

The second model we tested (Model 2) corresponds with the approach usually adopted by researchers who want to confirm a factor model: All high factor pattern coefficients were estimated as free parameters and all low factor pattern coefficients (i.e., pattern coefficients with $|f_{ij}| < .30$) were fixed to zero. Comparison of the results of Models 1 and 2 allows us to determine the extent to which a lack of fit in CFA is due to the more conservative use of CFA compared with EFA. More specifically, a factor model as specified in Model 2 (and as specified by most researchers who want to confirm a model) has many more restrictions than the factor model obtained through EFA. Therefore, we generally expected a bad fit of Model 2. It is important to note here that this would not automatically imply that CFA is too conservative or that EFA is too liberal. Rather, it points at differences in the use of the two techniques that may account for differences in results obtained by CFA and EFA.

Third, we tested a less constrained model (Model 3) in which we implemented some adjustments to Model 2. We freed or constrained some correlations between factors, and we freed some of the "secondary pattern coefficients" (i.e., $.20 < |f_{ij}| < .30$; Hofstee, De Raad, & Goldberg, 1992; Saucier & Goldberg, 1998) of the original exploratory solution in our confirmatory analysis. We only made these adjustments if they were likely to improve the fit as indicated by the Wald and Lagrange Multiplier (LM) tests (Bentler & Dijkstra, 1985; Lee, 1985). The Wald test indicates which of the free parameters may be dropped without seriously affecting the model fit. The LM test indicates which of the fixed or constrained parameters should be freed to improve the fit. These recommendations of the Wald and LM tests were implemented only if they did not lead to extreme deviations from the original exploratively obtained factor structure. In particular, we never allowed for correlations between error variances because correlated error terms are indications that the model has omitted one or more relevant exogenous variables, that is, factors (James, Mulaik, & Brett, 1982). Furthermore, factor pattern coefficients that were extremely low in the original exploratory factor structure ($|f_{ij}| < .20$) were never estimated as free parameters. However, factor pattern coefficients that were high in the original structure ($|f_{ij}| > .30$) and thus estimated as free parameters in Model 2 were free in Model 3 as well.

Obviously, we can expect Model 3 to have a better fit than Model 2 because Model 3 generally has fewer restrictions and the differences between the models are derived from tests aimed at improving the fit. Our main interest in Model 3 is, however, in the comparison of Model 3 with Model 1 (the most precise test of the original exploratory model). That is, in Model 3 we made some adjustments to Model 2 but only adjustments that did not lead to extreme deviations from the original exploratively obtained factor model. This is important because Model 3 is a "desimplification" of Model 2, and the desimplification of a factor model can easily lead to a decreased interpretability of the model.[2] However, by allowing only very moderate deviations from the original model, we aimed to develop a model that is interpretable in similar fashion as the original exploratory model. We were interested, therefore, in the extent to which results from Model 3 were comparable to results obtained in the direct test of the exploratively obtained factor structure, Model 1.

More specifically, if test statistics of Models 1 and 3 would be similar, this would imply that it might be justified to test an exploratory factor structure in a less stringent way than by Model 2. That is, it may be reasonable to release some of the secondary pattern coefficients when confirmatively testing a factor model. After all, although most of the individual secondary pattern coefficients frequently have little practical significance in that they account for little variance in a factor solution, together they form a large portion of the total variance. Indeed, some of the secondary pattern coefficients may frequently be statistically significant (for example, when sample size is large). Furthermore, by only freeing some of the secondary pattern coefficients, the substantive interpretation of the model is not appreciably altered. Therefore, it may sometimes be statistically necessary to unconstrain some of these secondary factor pattern coefficients when testing a factor model to evaluate the original factor structure.

In addition, it may be reasonable to allow for some correlations between factors, even if factors were orthogonal in the original exploratory solution. The reason for this is the following. By equating low factor pattern coefficients to zero, we formulate a model in which all correlations between variables that have their nonzero pattern coefficients on different factors are expected to be equal to zero. Obviously, this is a very restricted hypothesis that is very likely to be falsified because the low but nonzero pattern coefficients in the EFA results show that there must at least be some correlation between those variables. Allowing the factors to be correlated would ease those restrictions a little without freeing the zero factor pattern coefficients. On the other hand, this should be done with care because, as Thompson (1997) has shown, variables with zero pattern coefficients on correlated factors may still have considerable correlations with those factors. Therefore, freeing the orthogonality restriction of the factors may result in factors with a substantive interpretation that is very different from the one suggested by the zero pattern coefficients in the original, orthogonal solution.

Method

Material

In an electronic search through psychological research literature, we tried to find articles in which EFAs were described and that contained the original correlation matrices that were used in the EFAs. A set of approximately 100 potentially relevant articles was obtained. Articles that were actually available were explored and judged on relevance. An article was considered relevant when it contained (a) a matrix of factor pattern coefficients obtained by EFA and (b) a correlation matrix of the variables included in the analysis. Eventually, we selected the following articles: Strelau and Zawadski (1993); Yamauchi (1982); Alden, Wiggins, and Pincus (1990); Kremer (1990); Rice, Cole, and Lapsley (1990); and Davies, Stankov, and Roberts (1998). In these articles, a total of 10 factor structures are described, which vary with regard to the number of variables, number of factors, extraction method, and number of respondents. Furthermore, these articles all reported correlations of the variables included in the factor analyses to two decimal places. Although these articles did not constitute a random sample from all possible articles, we regard them as representative for the phenomena we want to study.

We will now briefly describe the 10 factor structures presented in the articles.[3] Furthermore, we will assign an individual number to each of the 10 factor structures (i.e., Structures 1 to 10). In the remainder of this article, we will refer to the factor structures by means of these numbers.

Description of the Factor Structures

In Strelau and Zawadski (1993), two different factor structures were distinguished in two different samples, resulting in four different structures. These four factor structures were all obtained by means of PCA with varimax rotations. In both Sample 1 and Sample 2 (N = 1,011 and 1,012, respectively), a factor structure with two orthogonal factors was extracted from five variables (we label these two Structures 1 and 2). Furthermore, in both Samples 1 and 2, a factor structure with four orthogonal factors was extracted from seven other variables (we label these two Structures 3 and 4). In Yamauchi (1982), a PCA with varimax rotation on eight variables (N = 124) resulted in a factor structure with four orthogonal factors (Structure 5). Alden et al. (1990) described a factor structure (obtained from a PCA, rotation method not reported) with two orthogonal factors, which were extracted from eight variables (N = 974; Structure 6). In Kremer (1990), a PCA with varimax rotation yielded a factor structure with five orthogonal factors (extracted from 14 variables) in a relatively small sample (N = 89; Structure 7). In Rice et al. (1990), an exploratory principal factor analysis with oblique rotation on eight variables (N = 120) yielded a factor structure with two slightly correlated factors (Structure 8). Finally, in Davies et al. (1998), two factor structures with their corresponding correlation matrices were described. Both factor structures were obtained by means of principal-axis EFA with oblimin rotation. In the first structure (Structure 9), eight correlated factors were extracted from 30 measures in a sample of 100 participants. In the second structure (Structure 10), five correlated factors were extracted from 17 measures in a sample of 131 participants.

Analysis

In the present study, we confirmatively examined all factor structures found in the aforementioned studies with the program EQS (Bentler, 1992). In the first model (Model 1), we fixed all factor pattern coefficients to their original values obtained in the exploratory analyses. In structures in which factors were correlated, the correlations were also fixed to their original values. The factor and error variances were added as free parameters. In the second model (Model 2), high factor pattern coefficients found in the exploratory model were specified as free parameters and low factor pattern coefficients in the exploratory model were fixed to zero in the CFA (a factor loading was considered high if $|f_{ij}| > .30$). If factors were orthogonal in the exploratively obtained structure, correlations between factors were fixed to zero. If factors were correlated in the exploratively obtained structure, correlations between factors were added as free parameters. To identify these models, the factor variances were fixed to one. In a few structures, it was necessary to constrain some other parameters to identify the model. More specifically, in Structures 3, 4, and 5, we had to fix one or two of the factor pattern coefficients to the values obtained in the exploratory factor structure to identify the model.

The third model tested (Model 3) was a less constrained model compared with Model 2. This model was obtained by following the recommendations of the Wald test and the LM test. The recommendations of the Wald and LM tests were implemented, as far as they did not lead to extreme deviations from the original exploratively obtained factor structure. That is, we never allowed for correlations between error variances as this would amount to adding new factors to the model (cf. James et al., 1982). Furthermore, factor pattern coefficients that were extremely low in the original exploratory factor structure ($|f_{ij}| < .20$) were never estimated as free parameters. Finally, pattern coefficients that were considered high enough to be estimated as free parameters in Model 2 ($|f_{ij}| > .30$) were always estimated as free parameters in Model 3 as well. Moreover, recommendations referring to estimating or constraining correlations between factors or estimating moderately low factor pattern coefficients ($.20 < |f_{ij}| < .30$) as free parameters were always implemented. With these adjustments, models were obtained that—in principle—would replicate the structure of the correlation matrices as closely as possible.

In all analyses, the maximum-likelihood method was used to estimate the parameters and to compute the fit of the model. The fit of a model was determined by first examining the significance of χ^2 in relation to the degrees of freedom. Given the problems of the χ^2 statistic alone (as described briefly in the introduction), we also examined two other fit indices: the NNFI (Bentler & Bonett, 1980) and the CFI (Bentler, 1990). Generally, the fit of a model is considered acceptable if these fit indices are equal to or larger than .90 (Pedhazur & Schmelkin, 1991).

Results

The results of our CFAs on the 10 exploratively obtained factor structures are described in Table 1. We will present the results produced by the three different models that were tested on the 10 factor structures.

Model 1: All Parameters Fixed to Their Original Values

When all parameters were fixed to their values obtained in the original exploratory factor structures, we found nonsignificant χ^2 values in four factor

Table 1

Summary of the Goodness-of-Fit of Three Models to Ten Data Sets

Factor Structure	Number of Variables	Number of Factors	N	Model 1 χ^2	(df)	Model 1 NNFI	Model 1 CFI	Model 2 χ^2	(df)	Model 2 NNFI	Model 2 CFI	Model 3 χ^2	(df)	Model 3 NNFI	Model 3 CFI
Structure 1	5	2	1,011	31.85	(8)***	0.97	0.98	32.09	(3)***	0.90	0.97	0.50	(1)	1.01	1.00
Structure 2	5	2	1,012	23.61	(8)**	0.98	0.98	160.40	(4)***	0.57	0.83	5.93	(1)*	0.95	1.00
Structure 3	7	4	1,011	158.30	(17)***	0.93	0.94	643.41	(14)***	0.61	0.74	89.76	(10)***	0.93	0.97
Structure 4	7	4	1,012	75.01	(17)***	0.97	0.98	621.96	(15)***	0.67	0.76	77.67	(10)***	0.95	0.97
Structure 5	8	4	124	1.51	(24)	1.19	1.00	48.37	(19)***	0.70	0.79	28.22	(17)*	0.87	0.92
Structure 6	8	2	974	460.70	(26)***	0.85	0.86	458.99	(16)***	0.75	0.86	440.11	(15)***	0.74	0.86
Structure 7	14	5	89	59.35	(86)	1.12	1.00	88.09	(70)	0.90	0.92	61.93	(68)	1.04	1.00
Structure 8	6	2	120	20.64	(13)	0.95	0.96	8.50	(6)	0.97	0.99	—		—	—
Structure 9	30	8	100	279.49	(427)	1.18	1.00	558.22	(391)***	0.77	0.80	493.89	(383)***	0.85	0.86
Structure 10	17	5	131	279.31	(131)***	0.68	0.69	220.68	(114)***	0.73	0.78	203.88	(112)***	0.77	0.81

Note. NNFI = nonnormed fit index; CFI = comparative fit index.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

structures (Structures 5, 7, and 9: *p*s > .98; Structure 8: $\chi^2[13, N = 120] = 20.64, p < .09$). The corresponding fit indices were all very high (all NNFIs > 0.94; all CFIs > 0.95), so these structures fitted very well. The χ^2 values of the six other factor structures were all highly significant (Structure 2: $\chi^2[8, N = 1{,}012] = 23.61, p < .01$; all other structures: *p*s < .001). However, the fit indices of Structures 1, 2, 3, and 4 were rather high (NNFIs > 0.92, CFIs > 0.93). Given that χ^2 alone is not always the most reliable indicator of model fit, the fit of these factor structures can be regarded as acceptable. Finally, Structures 6 and 10 had a bad fit (Structure 6: NNFI = 0.85, CFI = 0.86; Structure 10: NNFI = 0.68, CFI = 0.69).

In sum, most of the confirmatory models in which all parameters were fixed to the values that were found in the original exploratively obtained factor structures had a good fit. Out of 10 factor structures, 8 models fitted acceptably.

Model 2: High Pattern Coefficients Added as Free Parameters, Low Pattern Coefficients Fixed to Zero

In Model 2, high pattern coefficients were added as free parameters and lower pattern coefficients were fixed to zero. For factor structures that were based on correlated factors in the original exploratory structure, correlations between factors were also added as free parameters. As can be seen in Table 1, Structures 7 and 8 fitted very well: Both had a nonsignificant χ^2 (Structure 7: $\chi^2[70, N = 89] = 88.09, p < .08$; Structure 8: $\chi^2[6, N = 120] = 8.50, p < .21$) and acceptable fit indices (Structure 7: NNFI = 0.90, CFI = 0.92; Structure 8: NNFI = 0.95, CFI = 0.98). The other 8 structures all had very large and statistically significant χ^2 values (all *p*s < .001). Of these 8 structures, only Structure 1 had acceptable fit indices (Structure 1: NNFI = 0.90, CFI = 0.97). The 7 other factor structures all had very low fit indices (0.57 < all NNFIs < 0.78; 0.74 < all CFIs < 0.87). As expected, when high factor pattern coefficients were added as free parameters and lower factor pattern coefficients were fixed to zero, most factor structures (7 out of 10) could not be confirmed, even for the same data.

Model 3: A Less Constrained Model

Based on the LM and Wald tests described above, we then developed a less constrained model for each correlation matrix. It turned out that for Structure 8, the third model was identical to the second model, so we only tested a third model for the remaining nine factor structures. For most of the factor structures, some of the moderately low factor pattern coefficients ($.20 < |f_{ij}| < .30$) were added as free parameters. For most of the factor structures that were based on orthogonal exploratory factor structures, some correlations between factors were also added as free parameters. Nonsignificant χ^2 values were found for Structures 1 and 7 (ps > .48). These two structures had very high fit indices (Structure 1: NNFI = 1.01, CFI = 1.00; Structure 7: NNFI = 1.04, CFI = 1.00), which indicated that the structures fitted very well. The remaining seven factor structures had significant χ^2 values (Structures 2 and 5: ps < .05; Structures 3, 4, 6, 9, and 10: ps < .001). Three of these structures (Structures 2, 3, and 4) had acceptable fit indices (NNFIs > 0.91, CFIs > 0.94). Structure 5 had inconsistent fit indices (NNFI = 0.87, CFI = 0.92): Whereas the NNFI did not show an acceptable fit, the CFI did. Because the CFI showed an acceptable fit and the NNFI did approach 0.90, we regarded the fit of Structure 5 as marginally acceptable. Structures 6, 9, and 10 did not have acceptable fit indices (NNFIs < 0.86, CFIs < 0.87). For these structures, the less constrained model could not confirm the exploratively obtained factor structure adequately.

Thus, in six out of nine structures, the less constrained model fitted acceptably. When compared with Model 1, these six structures also fitted very well in Model 1, and two out of the three structures that did not fit acceptably had a bad fit in Model 1 as well. Apparently, the results obtained in the less constrained model seemed to be quite comparable to results from Model 1.

Discussion

The pattern of results revealed corroborative evidence for our line of reasoning. Model 1 had an acceptable fit in 8 out of 10 factor structures. Model 2, however, had an acceptable fit in only 3 out of 10 factor structures. As predicted, the fit of Model 2 on the same data was rather poor, both in absolute terms and relatively compared with Model 1. This lends support for the idea that the frequently found lack of fit in CFA may, at least to some extent, be attributable to the fact that CFA is typically applied with more restrictions than EFA. Model 3, on the other hand, had an acceptable fit in 6 out of 9 structures. If we compare these structures with Model 1, we note that most of the structures that fitted well in Model 1 also fitted well in Model 3, whereas the 2 structures that fitted badly in Model 1 also fitted badly in Model 3.[4] Thus, Model 3 seems to be quite accurate in replicating the findings of Model 1. This is an indication that it may be justified to implement such a less constrained model when confirming a factor model, as we will discuss below.

Although our study focused on the role of lower factor pattern coefficients, our results do not imply that the restriction of low and nonsignificant pattern coefficients is the only reason why an exploratively obtained factor structure often cannot be confirmed. Other explanations may also account for this to some extent. First of all, in EFAs, different researchers may have used different criteria to define the factor solution (Gorsuch, 1983; cf. Vassend & Skrondal, 1997). This could mean that in some cases, the optimal solution for the data at hand might not have been discovered in the exploratory analysis

(Fabrigar et al., 1999). As a result, the exploratively obtained factor structure may sometimes yield a bad starting point for confirmative analyses. Furthermore, as noted in the introduction, conceptually unrelated technical conditions like sample size (e.g., Boomsma, 1987) and deviations from multivariate normality (e.g., Hu et al., 1992) may also affect the CFA model fit.

It is important to note that we do not assume that in the exploratory analysis of our data the optimal solution was found or that sample size or the multivariate normality assumption was optimal. Indeed, violations of these assumptions may possibly explain why a good fit could not be obtained in two factor structures when we specified the exact exploratively obtained factor structure (i.e., Model 1). Moreover, even when a good fit was obtained in Model 1, the extent to which these assumptions were met may not always have been optimal. Yet, it should be recognized that the extent to which these assumptions were optimal was (within structures) equal between Models 1, 2, and 3. After all, Models 1, 2, and 3 were based on the same exploratively obtained factor structures and on the same data. Therefore, even if these assumptions were not optimal in a factor structure, they cannot account for differences in results between Models 1, 2, and 3.

Practical Implications

How should researchers proceed when they want to confirm an exploratively obtained factor structure? Based on the preliminary findings here, one could argue that researchers should start out to test Model 1. After all, this is the closest replication of the original factor structure, which, generally, turns out to fit reasonably when applied to the same data. However, from a practical point of view, things are often not that simple. That is, one may often find that when an EFA is reported in the literature, the article only shows the higher pattern coefficients and does not report the lower pattern coefficients. In fact, this was one of the problems we encountered during the search process for exploratively obtained factor structures that were reported in the literature.

Researchers may therefore want to consider developing a somewhat less restricted model (our Model 3). As noted above, Model 3 seemed to be a rather close replication of Model 1. Model 3 was computed by allowing for some correlations between factors and by freeing some of the secondary factor pattern coefficients (.20 $|f_{ij}| < .30$). This may have some important practical implications for researchers who want to confirm an exploratively obtained factor model. First, variables that have a factor loading fixed to zero may in reality be correlated with a factor. Deleting this variable-factor correlation may result in factors that inevitably cannot explain all correlations between the variables. By way of compensation, it seems reasonable, therefore, to allow for some correlations between factors in a confirmatory factor model, even though the model may be based on an exploratively obtained structure with orthogonal factors. However, care should be taken that the variables with zero pattern coefficients should not obtain too large correlations (i.e., structure coefficients) with the factors. Otherwise, one would be testing a model that is substantively different from the model one started out to test (cf. Thompson, 1997).

Second, in our study, freeing the secondary factor pattern coefficients that are indicated by the LM test was found to compensate somewhat for the restriction of all low factor pattern coefficients to zero. That is, secondary pattern coefficients form a substantial amount of the total variance and may even be statistically significant in a factor model. Furthermore, these pattern coefficients were free parameters in the original EFA analysis. As a consequence, a model in which some of these pattern coefficients are freed would be more accurate in replicating the original exploratively obtained model. Freeing some of the secondary pattern coefficients does not decrease the interpretability of the model substantially because the adjustments of the model are within the results of the original exploratory model. It seems justified, therefore, to free some of the secondary factor pattern coefficients when testing a factor model by CFA.

In closing, we would like to remark that the above considerations could and indeed need to be complemented by statistical research regarding the shrinkage of factor models when applied to new data sets. If a good fit is questionable when the factor structure is confirmatively tested on the same data, we cannot expect that a test of the factor structure in a confirmative follow-up study, that is, on different data, will lead to a good fit.

Notes

1. Note that there is never one single true model. Even if there were a model that perfectly reproduced the original correlation matrix, any rotation of the model would also fit the data perfectly.

2. We thank an anonymous reviewer for urging us to emphasize this point.

3. Due to space considerations, we decided not to print tables with all of the factor structures and correlation matrices. The factor structures and corresponding correlation matrices on which we based our analyses were published in the articles referred to or can be found at http:www.fsw.leidenuniv.nl/www/w3_ment/medewerkers/vdkloot/vdklootcv.htm.

4. The only exception is Structure 9, in which Model 1 fitted very well, whereas Model 3 fitted less than acceptable. This may be explained if we take into account that Structure 9 was a rather large factor model (i.e., a structure with eight factors extracted out of 30 variables). After freeing some of the moderately low factor pattern coefficients in such a large model, many of the lower factor pattern coefficients may still be fixed to zero. Thus, in Structure 9, a large number of the parameters may have been restricted in Model 3, even after freeing some of the moderately low factor pattern coefficients.

References

Alden, L. E., Wiggins, J. S., & Pincus, A. L. (1990). Construction of circumplex scales for the inventory of interpersonal problems. *Journal of Personality Assessment*, *55*, 521-536.

Bandalos, D. L. (1993). Factors influencing cross-validation of confirmatory factor analysis models. *Multivariate Behavioral Research*, *28*, 351-374.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246.

Bentler, P. M. (1992). *EQS: structural equations program manual*. Los Angeles: BMDP Statistical Software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588-606.

Bentler, P. M., & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P. R. Krishnaiah (Ed.), *Multivariate analysis VI* (pp. 9-42). Amsterdam, the Netherlands: Elsevier.

Bollen, K. A. (1986). Sample size and Bentler and Bonnett's nonnormed fit index. *Psychometrika*, *51*, 375-377.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.

Bollen, K. A. (1990). Overall fit in covariance structure models: Two types of sample size effects. *Psychological Bulletin*, *107*, 256-259.

Boomsma, A. (1987). The robustness of maximum likelihood estimation in structural equation models. In P. Cuttance & R. Ecob (Eds.), *Structural modeling by example: Applications in educational, sociological, and behavioral research* (pp. 160-180). New York: Cambridge University Press.

Borkenau, P., & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study of the 5-factor model of personality. *Personality and Individual Differences*, *11*, 515-524.

Byrne, B. M. (1989). *A primer of LISREL: Basic applications and programming for confirmatory factor analytic models*. New York: Springer-Verlag.

Church, A. T., & Burke, P. T. (1994). Exploratory and confirmatory tests of the big five and Tellegen's three- and four-dimensional models. *Journal of Personality and Social Psychology*, *66*, 93-114.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16-29.

Davies, M., Stankov, L., & Roberts, R. D. (1998). Emotional intelligence: In search of an elusive construct. *Journal of Personality and Social Psychology*, *75*, 989-1015.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272-299.

Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale, NJ: Lawrence Erlbaum.

Grice, J. W., & Harris, R. J. (1998). A comparison of regression and loading weights for the computation of factor scores. *Multivariate Behavioral Research*, *33*, 221-247.

Hartman, C. A., Hox, J., Auerbach, J., Erol, N., Fonseca, A. C., Mellenbergh, G. J., Nøvik, T. S., Oosterlaan, J., Roussos, A. C., Shalev, R. S., & Zilber, N. (1999). Syndrome dimensions of the Child Behavior Checklist and the Teacher Report Form: A critical empirical evaluation. *Journal of Child Psychology and Psychiatry*, *40*, 1095-1116.

Hofstee, W.K.B., De Raad, B., & Goldberg, L. R. (1992). Integration of the big five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology*, *63*, 146-163.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424-453.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*, 351-362.

James, L. R., Mulaik, S. A., & Brett, J. (1982). *Causal analysis: Models, assumptions and data*. Beverly Hills, CA: Sage.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and application*. Chicago: SPSS.

Kremer, J. F. (1990). Construct validity of multiple measures in teaching, research, and service and reliability of peer ratings. *Journal of Educational Psychology*, *82*, 213-218.

Lee, S. Y. (1985). On testing functional constraints in structural equation modeling. *Biometrika*, *72*, 125-131.

Lonigan, C. J., Hooe, E. S., David, C. F., & Kistner, J. (1999). Positive and negative affectivity in children: Confirmatory factor analysis of a two-factor model and its relation to symptoms of anxiety and depression. *Journal of Consulting and Clinical Psychology*, *67*, 374-386.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*, 391-410.

McCrae, R. R., Zonderman, A. B., Costa, P. T., Jr., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the revised NEO Personality Inventory: Confirmatory factor analysis versus procrustes rotation. *Journal of Personality and Social Psychology*, *70*, 552-566.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, *107*, 247-255.

Parker, J.D.A., Bagby, R. M., & Summerfeldt, L. J. (1993). Confirmatory factor analysis of the revised NEO Personality Inventory. *Personality and Individual Differences*, *15*, 463-466.

Pedhazur, E., & Schmelkin, L. P. (1991). *Measurement, design and analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum.

Rao, N., & Sachs, J. S. (1999). Confirmatory factor analysis of the Chinese version of the Motivated Strategies for Learning Questionnaire. *Educational and Psychological Measurement*, *59*, 1016-1029.

Reynolds, A. J., & Lee, J. S. (1991). Factor analyses of measures of home environment. *Educational and Psychological Measurement*, *51*, 181-192.

Rice, K. G., Cole, D. A., & Lapsley, D. K. (1990). Separation-individuation, family cohesion, and adjustment to college: Measurement validation and test of a theoretical model. *Journal of Counseling Psychology*, *37*, 195-202.

Saucier, G., & Goldberg, L. R. (1998). What is beyond the big five? *Journal of Personality*, *66*, 495-524.

Strelau, J., & Zawadski, B. (1993). The formal characteristics of behaviour-temperament inventory (FCB-TI): Theoretical assumptions and scale construction. *European Journal of Personality*, *7*, 313-336.

Thompson, B. (1997). The importance of structure coefficients in structural equation modeling confirmatory factor analysis. *Educational and Psychological Measurement*, *57*, 5-19.

Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.

Vassend, O., & Skrondal, A. (1997). Validation of the NEO Personality Inventory and the five-factor model. Can findings from exploratory and confirmatory factor analysis be reconciled? *European Journal of Personality*, *11*, 147-166.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Issues, concepts, and applications* (pp. 56-75). Thousand Oaks, CA: Sage.

Yamauchi, H. (1982). Sex differences in motive to avoid success on competitive or cooperative action. *Psychological Reports*, *50*, 55-61.